Final Project: Intro to R for Epidemiologists

Requirements

- Submit an R markdown file with your code and analysis named "LASTNAME_560R_project.Rmd"
- Your R markdown file should only print the required elements (see below) and should not print any code.
- **You may ask the teaching assistants and instructor for help, but you may not consult with other students about any aspect of the project.**
- Submit your project to jenna.krall@emory.edu by 5 PM on Thursday 4/30/15

Data

Option 1. Provide your own data

You may use your own dataset for the final project. Your data:

- Must have at least 6 variables (with at least 2 continuous variables)
- Must have at least 100 observations
- Must be able to answer a relevant scientific question (e.g. is air pollution associated with mortality?)
- You will not have to share your data with the instructor or TAs if there are restrictions on who may access the data (e.g. data use agreements)
- **You must have your dataset approved by the instructor by Thursday, March 26**

Option 2. Use provided data

Investigate whether smoking is associated with coronary heart disease using the dataset `project560Rdata.csv` on the course website. The data are from a 9 year prospective study of white males in Georgia. The variables include:

id	Subject identifier
chd	Coronary heart disease (1 = yes)
cat	Normal catecholamine level (1 = yes)
age	Age in years
chl	Cholesterol
smk	Ever smoked (1 = yes)
ecg	Electrocardiogram abnormality (1 = yes)
dbp	Diastolic blood pressure
sbp	Systolic blood pressure
hpt	High blood pressure (1 = yes)

Project details

Your final report should have the following sections:

1. Exploratory data analysis (20 points)
   (a) One table of summary statistics for each variable
   (b) One figure displaying relevant exploratory information (**using ggplot2**)
2. Univariate data analysis (e.g. Is your outcome associated with each relevant predictor? You may use t-tests, chi-squared tests, or regression analysis) (15 points)
   (a) One table of univariate test results
3. Multivariate data analysis (i.e. estimate associations between relevant covariates and your outcome using regression models with more than one covariate) (25 points)
   (a) One table of multivariate regression results
   (b) One figure displaying relevant multivariate results (**using either base plotting or ggplot2**)
4. Summary paragraph (10 points)
   (a) In one paragraph, summarize your findings. Include relevant statistics from 1-3.

To receive full credit, you must:

- Hide your code so that your final report only contains the tables, figures, and summary paragraph described above.
- Caption all tables and figures.